LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-CONF-209117

New Approaches for Collaborative Sharing of Chemical Model Data and Analysis Tools

K. Schuchardt, O. Oluwole, W. Pitz, L. A. Rahn, W. H. Green, Jr., D. Leahy, C. Pancerella, M. Sjöberg, J. Dec

January 20, 2005

2005 Joint Meeting of the U.S. Sections of The Combustion Institute
Philadelphia, PA, United States
March 20, 2005 through March 23, 2005

New Approaches for Collaborative Sharing of Chemical Model Data and Analysis Tools

Karen Schuchardt,[1] Oluwayemisi Oluwole,[2] William Pitz,[3] Larry A. Rahn,[4] William H. Green, Jr.,[2] David Leahy,[4] Carmen Pancerella,[4] Magnus Sjöberg,[4] John Dec[4]

[1]*Pacific Northwest National Laboratory, Richland, WA 99352*
[2]*MIT, Cambridge, MA 02139*
[3]*Lawrence Livermore National Laboratory, Livermore, CA 94551*
[4]*Sandia National Laboratories, Livermore, CA 94551*

Introduction

The urgent need for high-efficiency, low-emission energy utilization technologies for transportation, power generation, and manufacturing processes presents difficult challenges to the combustion research community. The required predictive understanding requires systematic knowledge across the full range of physical scales involved in combustion processes – from the properties and interactions of individual molecules to the dynamics and products of turbulent multi-phase reacting flows. Innovative experimental techniques and computational approaches are revolutionizing the rate at which chemical science research can produce the new information necessary to advance our combustion knowledge. But the increased volume and complexity of this information often makes it even more difficult to derive the systems-level knowledge we need. Combustion researchers have responded by forming interdisciplinary communities intent on sharing information and coordinating research priorities. Such efforts face many barriers, however, including lack of data accessibility and interoperability, missing metadata and pedigree information, efficient approaches for sharing data and analysis tools, and the challenges of working together across geography, disciplines, and a very diverse spectrum of applications and funding.

This challenge is especially difficult for those developing, sharing and/or using detailed chemical models of combustion to treat the oxidation of practical fuels. This is a very complex problem, and the development of new chemistry models requires a series of steps that involve acquiring and keeping track of a large amount of data and its pedigree. Also, this data is developed using a diverse range of codes and experiments spanning *ab initio* chemistry codes, laboratory kinetics and flame experiments, all the way to reacting flow simulations on massively parallel computers. Each of these processes typically requires different data formats, and often the data and/or analysis codes are only accessible by personally contacting the creator. Chemical models are usually shared in a legacy file format, such as Chemkin [1] or FlameMaster [2], often without needed metadata and pedigree information. Detailed reaction mechanisms are usually too large for efficient operation of chemical reacting flow solvers so there is also much work aimed at reducing number of species and reactions in a fashion consistent with the accuracy needs of the simulation. In the best of cases, this results in a proliferation of diverse models that are difficult to find or trace to their origins.

The Collaboratory for Multi-scale Chemical Science (CMCS) brings together leaders in scientific research and technological development across multiple DOE laboratories, other government laboratories and academic institutions to develop an informatics-based approach to synthesizing multi-scale information to create knowledge in the chemical sciences. CMCS applies advanced collaboration and metadata-based data management technologies to develop an open source multi-scale informatics toolkit serving as the basis for a CMCS web portal. The portal enables cross-scale data discovery, viewing, comparison, transformation and exchange while facilitating community formation, communication, and data development. The portal also includes tools for browsing cross-scale data dependencies and mechanisms to integrate custom and community resources into active research projects. CMCS researchers are populating the portal with key chemistry data and integrating important chemistry applications. Additionally, several new chemical informatics applications supported by the project are being made available as portal-centric web services. These capabilities are being piloted internationally across several scales by interacting groups of leading researchers in combustion science.

Of particular interest here, is that the CMCS team is developing new approaches that facilitate collaborative sharing of chemical model data and analysis codes. In this paper, we will briefly describe the relevant CMCS infrastructure and tools in the context of an ongoing collaboration motivated by the desire to understand and control Homogenous Charge Compression Ignition (HCCI) in internal combustion engines. We will relate the issues and progress in developing XML based formats for large detailed chemical models [3], methods for documentation and viewing metadata, and implementations of translations to other useful formats. As an example of one way collaborative analysis codes can be implemented in CMCS, we will focus on the Range Identification and Optimization Tool (RIOT). RIOT is a software tool designed to reduce the number of reactions in a reaction mechanism to just those needed to maintain a specified chemical accuracy in the predictions of the resulting reduced mechanism [4, 5]. The approach for the ultimate use of such reduced models in HCCI research will also be briefly described. We intend to present in this way a vignette of how CMCS is providing new approaches to make research with practical combustion models easier, and to invite broader participation.

HCCI Combustion Challenge

Homogeneous charge, compression ignition is a piston-engine combustion process that has emerged as a high-efficiency alternative to traditional spark-ignition (SI) combustion and a low-emissions alternative to traditional diesel combustion. It

offers good fuel economy, similar to a diesel engine, while producing very low emissions of NO_X and soot particulate matter (PM). However, before HCCI can be implemented in production engines, several technical barriers must be overcome. Therefore, research and development is being pursued in a number of areas, including: control of ignition timing over the load/speed map, slowing the heat-release rate at higher loads, controlling hydrocarbon (HC) and carbon monoxide (CO) emissions, maintaining combustion stability, cold starting, and response to load and speed transients.

Computer modeling of the in-cylinder processes in an HCCI engine is a key element in this research and development effort. Modeling methodologies for HCCI range from adiabatic single-zone models [6],[7] to integrated multidimensional CFD/kinetics models [8],[9] engine models. Due to the nature of HCCI, the ignition, combustion heat-release, and pollutant formation processes are much more highly coupled to the chemical-kinetics of the charge mixture than they are for traditional SI or diesel combustion. Therefore, the usefulness of HCCI modeling results relies heavily on the accuracy of the chemical-kinetics mechanism employed and their ability to match the behavior of the specific fuel in use. For a limited number of reference fuels, detailed chemical mechanisms have been developed [10],[11],[3]. These mechanisms often contain many hundreds of species and thousands of reactions, and have provided useful understanding of the in-cylinder processes [6],[8],[12]. Yet, even these detailed mechanisms sometimes fail to describe important changes to the autoignition with changes in engine operation conditions [7],[13]. Thus, there is a strong need for improvement and further development of these detailed chemical kinetics mechanisms. Also, because real fuels (such as gasoline) are combinations of many compounds, there is a need to develop chemical-kinetic models for surrogate mixtures that accurately reproduce the behavior of real fuels [13].

Improved kinetic mechanisms would substantially improve our ability to simulate HCCI combustion, and therefore, provide a valuable component in overcoming the technical barriers to HCCI. For example, improved detailed mechanisms used in single-zone modeling would help to enhance the understanding of how the autoignition process is affected by various engine parameters. This is important for effective control of the ignition timing as engine operating conditions change. On the other hand, a combination of multi-dimensional CFD and chemical-kinetics modeling is required to understand the effects of non-uniformities in the charge mixture and temperature. This is particularly important for investigations of techniques to slow the heat-release rate to allow operation at higher loads (such as the use of non-uniform in-cylinder temperature fields), or for understanding the details of the in-cylinder combustion and emissions formation processes. For such models, the computational times can quickly become impractical if detailed chemical kinetics mechanisms are used. Therefore, it is necessary to reduce the detailed mechanisms to a more tractable size without sacrificing the ability to capture the important chemistry of the conditions studied. The CMCS provides a valuable tool for facilitating the development and tuning of these required chemical-kinetic mechanisms and their reduced-mechanism counterparts.

CMCS Infrastructure and Tools Being Developed to Meet the HCCI Challenge

CMCS is developing open-sourced software called KnECS (Knowlege Environment for Collaborative Science) that offers many features to meet the challenge of tracking large amounts of data and its pedigree while fostering sharing and data development. KnECS is a multi-scale informatics toolkit that addresses issues related to knowledge sharing such has provenance tracking, federation of data, and application resources. The KnECS toolkit has a multi-layer architecture including a web portal interface and underlying services that together provide rich group-level collaboration capabilities, data/metadata organization and management capabilities, and multiple interfaces for integrating third-party capabilities such as domain specific data viewers and analysis tools, data translators, and experimental or computational data sources. The portal engine is based on CompreHensive collaborativE Framework (CHEF) [14] which uses Apache Jetspeed [15], but has been modified to meet CMCS needs. Many of the software features are enabled by the KnECS data infrastructure which is provided by Scientific Annotation Middleware (SAM) [16]. More about the vision of CMCS and its informatics infrastructure can be found a recent paper [17] and on the CMCS web site (cmcs.org).

As mentioned, KnECS and CMCS provide many capabilities and mechanisms by which the capabilities may be accessed. CMCS supports the integration of scientific applications through portlet interfaces that access scientific codes through web services. Examples include Active Thermochemical Tables (ATcT) [18] and RIOT [4] with others planned. CMCS is also developing tools centered around the community data sharing and curration processes. Additionally, CMCS has implemented numerous translations for various chemical data formats. Desktop applications can also directly access the KnECS infrastructure using the Data Service Interface (DSI) library that provides a java interface to the data middleware. Finally, non-CMCS desktop applications can access data using the WebDAV protocol or through clients that are available with or for many operating systems [19-22].

Meeting the Informatics Challenges of Detailed Chemical Models

The development of chemical kinetic models offers many challenges that can be overcome with informatics technologies. To construct a chemical kinetic model, thermodynamic data must be obtained for each species relevant the oxidation of the fuel. The thermodynamic data can be obtained from the literature, from *ab initio* chemistry calculations, by group additivity and other methods. Active Thermochemical Tables (ATcT) can be used to find the most accurate estimate of enthalpy of formation for a species [18]. Next, a series of reactions must be formulated to address the oxidation of the fuel. Rate constants for each reaction must be obtained from the literature or by estimation techniques. Pedigree, such as the

origin of rate constants and thermodynamic data, must be created, stored and made accessible to scientific investigators. The reaction mechanism also must be validated by comparing predicted results with experimental results for the fuel of interest. The reaction mechanism is shared with collaborators who often need to be able to follow the pedigree of the rate constants, thermodynamic, and validation data.

The CMCS and KnECS offer ways to help make development of detailed chemical mechanisms easier. With this informatics approach, users can track the development of detailed chemical kinetic models all the way from *ab initio* calculations of species and reaction rate constants to development of reduced chemical kinetic models for reacting flow codes. With this informatics approach, for example, a scientist can access a reduced model, find the detailed chemical kinetic mechanism from which it was derived, find the conditions for which it is valid, and, if desired, follow the pedigree back to actual rate constant estimates and the *ab initio* chemistry calculations on which they are based.

This informatics approach is enabled by CMCS in several ways. First, the data is stored in an XML format so that metadata such as rate-constant references and pedigree can be stored and translations of the data into legacy formats and html views can easily be made available. CMCS has developed XML-based formats for large detailed models and for species thermodynamic data. Secondly, applications such as electronic structure calculations, group additivity calculations for thermodynamics, ATcT estimates for thermodynamic properties, and RIOT calculations for mechanism reduction, are enabled through the CMCS portal so that pedigree data can be kept track of, and data can be stored in XML formats so that metadata are available. Examples of metadata are the version of a code that was used to do a calculation, and a description of how a chemical kinetic rate constant was estimated. Also, XML formatted data allows easy translation of data to legacy formats such as Chemkin [1] and to enable convenient viewing, for example in plots, html tables and molecular views [23].

The CMCS portal facilitates the sharing of data so that scientists can work together more effectively to develop and use chemical kinetic mechanisms. Users can form groups whose membership is controlled by its members. Groups can share of scientific data through the portal using secure SSL technology. The group can access shared files, be notified when new data is available, search the data warehouse, use a group calendar, set up group tasks, and use discussion threads and chat.

RIOT Software

RIOT (Range Identification and Optimization Toolkit) is a software tool to reduce the number of species and/or reactions in a reaction mechanism while maintaining user specified tolerances on the accuracy of the reduced mechanism [4, 5]. The basic idea is to reduce the model without significantly changing predicted values such as temperature, pressure and species concentrations from the values predicted by using the full model. The computational gain realized by using the reduced models is proportional to the extent of the reduction. Therefore, maximum speedup can be obtained by using optimally-reduced models. Detailed chemical kinetics mechanisms for practical fuels are too large to be accommodated by multidimensional, chemically reacting flow codes. These mechanisms often contain thousands of species and reactions which are much more than can be handled by present computational resources. Reaction mechanism reduction is usually needed. CMCS has integrated the RIOT model reduction code into its portal environment by developing a portlet interface that assists the user with specifying inputs to the RIOT code, generates the input file, ships the request to a web service interface to the code, and uploads and presents the results when the reduction completes. We will now describe how this capability can be used through a practical example being worked on by participating CMCS scientists.

Example Case: Reduction of a Mechanism for Use with HCCI

The reacting flow modeling of HCCI requires the use of detailed chemical kinetic mechanisms because ignition in the engine is controlled by the homogeneous ignition of the fuel-air mixture. However, chemical kinetic models to simulate gasoline and diesel fuel contain thousands of species and reactions and need to be reduced to obtain shorter run times for the reacting flow codes. In the example given below, n-heptane is used as a surrogate fuel for diesel fuel. N-heptane, which has a Cetane number of 55, has similar ignition quality as diesel fuel. Cetane number is a measure of diesel fuel ignitability. In the following example, a detailed chemical kinetic mechanism for n-heptane [3, 11] is reduced using the RIOT portlet. This detailed mechanism has 2539 reactions.

Using the RIOT Portal to Reduce a Mechanism

We now describe the process of using the RIOT portlet to reduce a detailed chemical kinetic mechanism. The RIOT program requires several input files to perform the mechanism reduction. These files include a detailed chemical kinetic mechanism file, a thermodynamics data file and one or more solution files. The first two files are assumed to be in Chemkin format [1]. The solution files contain histories of pressure, temperature, and species concentrations. For HCCI, this would likely be computed histories from a simulation of an HCCI ignition in an engine. In addition to a solution file format specified by RIOT, the portal accepts both Senkin and Aurora [1] solution files from their post processors to provide these histories. The portlet automatically recognizes the file formats (mime type) if the file names end with the extension ".senkin-export" or ".aurora-export". This is an example where we take advantage of the portal capability of recognizing file mime types and enabling data translations. To support additional file formats, new translators can be written and registered with the CMCS infrastructure. Without changes to the RIOT portlet itself, these new formats will automatically be recognized by the RIOT portlet.

Once the user has assembled the required files, the user uploads them to the portal using the Data Browser portlet. This portlet is the CMCS data and metadata browsing tool providing a full set of basic content management operations including file upload, copy, move, and delete, and the creation of hierarchical directory structures [17]. The portal recognizes the reaction mechanism file and the thermodynamic data file based on their mime-types (".mech" and ".therm" respectively). Translations to CMCS standard XML-based formats are automatically available to the user enabling further translations through XSLT scripts such as a table view of a mechanism (shown later). The user then accesses the RIOT portlet and specifies the set of files to be used for the reduction. One or more solutions files can be specified. Multiple solution files may be needed to cover a broad range of conditions (pressure, temperature, fuel composition, equivalence ratio and exhaust gas recirculation) for HCCI. Then the user is presented with the main page of the RIOT portlet (Fig. 1).

In the RIOT portlet shown in Fig. 1, the user selects the solution points to be used to reduce the model and specifies the tolerances that affect the accuracy of the model. The solution points are selected under the window labeled "Reaction Conditions" where the available points and conditions are shown in a list. The user selects the points that have the critical temperatures and pressures where the reduced mechanism should be valid. In this case, four solution points are selected. Then "Model Reduction Parameters" button provides access to a page where parameters for which the reduced model should be accurate are specified including tolerance in temperature, species concentrations, and even which reactions should never be removed (Fig. 3). The last line on the screen shot of Fig. 1 gives the total number of species (561) and reactions (2539) in the detailed chemical kinetic mechanism. The user then clicks on "Submit" to reduce the mechanism. The mechanism and all required inputs are sent via an asynchronous web-based service to a server at MIT where the mechanism reduction is performed. When the calculation is finished, an xml-based reduced mechanism file is returned to the CMCS data store where it is available in the RIOT portlet. When the entire process completes, the user receives an email notification with a link that leads directly to the file in the portal. The xml-formatted results file allows easy XSLT-based automatic conversion to a table view (Fig. 2) or to a Chemkin formatted file for use with desired reacting flow calculations. As described above for input files, additional translation scripts can be written and registered with CMCS. For example, translation to other reacting flow code mechanism formats can be added.

Another illustration of the capabilities of CMCS is its ability to follow the pedigree of the data. In Fig. 4, we see that the portal provides a graph that allows a user to track the pedigree of the reduced model. The user can trace the reduced model back to the detailed model and the specified conditions that were used to create it. It is part of the CMCS vision that users will be able to follow the pedigree of the data without having to contact other scientists who created the data. This feature should help to reduce the time it takes for scientists to understand where data came from and what its limitations are. The resulting reduced mechanism in Fig. 2 has 73 reactions (too many reactions to show). This reduced mechanism was obtained using more than the four solution points shown in Fig. 1. This example of a reduced mechanism has not been validated but it includes many reactions that would be expected to be important based on previous studies of HCCI ignition.

Acknowledgments

The authors wish to thank the team members of the Collaboratory for Multi-Scale Chemical Science (CMCS) for their many contributions to this work. This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Science National Collaboratory Program. The authors thank program manager Mary Anne Scott for her support of this work. This work was performed under the auspices of the U.S. Department of Energy by UC LLNL under contract No. W-7405-Eng-48, Sandia National Laboratories under Contract No. DE-AC04-94-AL85000, and Pacific North Pacific Northwest National Laboratory under Contract No. DE-AC06-76RLO 1830.

References

1. Kee, R. J., Rupley, F.M., Miller, J.A., Coltrin, M.E., Grcar, J.F., Meeks, E., Moffat, H.K., Lutz, A.E., Dixon-Lewis, G., Smooke, M.D., Warnatz, J., Evans, G.H., Larson, R.S., Mitchell, R.E., Petzold, L.R., Reynolds, W.C., Caracotsios, M., Stewart, W.E., Glarborg, P., Wang, C., Adigun, O., Houf, W.G., Chou, C.P., Miller, S.F. *Chemkin Collection*, 3.7.1, Reaction Design, Inc., San Diego, CA, 2003.
2. Pitsch, H., Entwicklung eines Programmpaketes zur Berechnung eindimensionaler Flammen am Beispiel einer Gegenstromdiffusionsflamme, Master's thesis, RWTH Aachen Germany, Aachen, 1993.
3. Pitz, W. J., Westbrook, C. K. and Curran, H. J. *LLNL Chemical Kinetic Mechanisms*, 2004.
4. Bhattacharjee, B., Schwer, D. A., Barton, P. I. and Green, W. H., *Combust. Flame* 135(3):191-208 (2003).
5. Olumwole, O. O., Bhattacharjee, B., Tolsma, J. E., Barton, P. I. and William H. Green, J., *Combust. Flame,* submitted, 2005.
6. Sjöberg, M. and Dec, J. E., "An Investigation into Lowest Acceptable Combustion Temperatures for Hydrocarbon Fuel in HCCI Engines," *Proc. Combust. Inst.* Vol. 30 pp. 2719 - 2726 (2004).
7. Dec, J. E. and Sjöberg, M., *Isolating the Effects of Fuel Chemistry on Combustion Phasing in an HCCI Engine and the Potential of Fuel Stratification for Ignition Control*, SAE Paper 2004-01-0557 (2004).

8. Aceves, S. M., Flowers, D. L., Espinosa-Loza, F., Martinez-Frias, J., Dec, J. E., Sjöberg, M., R.W. and Hessel, R. P., *Spatial Analysis of Emissions Sources for HCCI Combustion at Low Loads Using a Multi-Zone Model*, SAE Paper 2004-01-1910, (2004).
9. Embouazza, M., Haworth, D. C. and Darabiha, N., *Implementation of Detailed Chemical Mechanisms Into Multidimensional CFD Using in Situ Adaptive Tabulation: Application to HCCI Engines*, SAE Paper 2002-01-2773, (2002).
10. Curran, H. J., Gaffuri, P., Pitz, W. J. and Westbrook, C. K., *Combust. Flame* 129:253-280 (2002).
11. Curran, H. J., Gaffuri, P., Pitz, W. J. and Westbrook, C. K., *Combust. Flame* 114(1-2):149-177 (1998).
12. Sjöberg, M., Dec, J. E. and Cernansky, N. P., *Potential of Thermal Stratification and Combustion Retard for Reducing Pressure-Rise Rates in HCCI Engines, based on Multi-Zone Modeling and Experiments*, SAE Paper 2005-01-0113, (2005).
13. Naik, C. V., W.J., P., Sjöberg, M., J.E., D., Orme J., C., H.J., J.M., S. and Westbrook, C. K., "Detailed Chemical Kinetic Modeling of Surrogate Fuels for Gasoline and Application to an HCCI Engine," *2005 Joint Meeting of the U.S. Sections of The Combustion Institute*, 2005.
14. CHEF, *Collaborative Portal Framework Website, CHEF Collaborative Portal Framework Website, http://www.chefproject.org/, University of Michigan*, University of Michigan, (2004).
15. Jetspeed, *an Open Source implementation of an Enterprise Information Portal, using Java and XML, Apache Jakarta Project, http://jakarta.apache.org/jetspeed/site/*, (2003).
16. Myers, J. D., Chappell, A., M., E., A., G. and Schwidder, J., *IEEE Computing in Science and Engineering* 5(3):44-50 (2003).
17. Myers, J. D., Allison, T. C., Bittner, S., Didier, B., Frenklach, M., William H. Green, J., Ho, Y.-L., Hewson, J., Koegler, W., Lansing, C., Leahy, D., Lee, M., McCoy, R., Minkoff, M., Nijsure, S., Laszewski, G. v., Montoya, D., Pancerella, C., Pinzon, R., Pitz, W., Rahn, L. A., Ruscic, B., Schuchardt, K., Stephan, E., Wagner, A., Windus, T. and Yang, C., "A Collaborative Informatics Infrastructure for Multi-scale Science," *Challenges of Large Applications in Distributed Environments (CLADE 04)*, Honolulu, Hawaii, 2004, p. 24, http://csdl.computer.org/comp/proceedings/clade/2004/2115/00/2115toc.htm.
18. Ruscic, B., Pinzon, R. E., Morton, M. L., von Laszevski, G., Bittner, S. J., Nijsure, S. G., Amin, K. A., Minkoff, M. and Wagner, A. F., *J. Phys. Chem. A* 108(45):9979-9997 (2004).
19. WebDrive *http://www.webdrive.com/*, South River Technologies: 2004.
20. Davfs *WebDAV Linux File System, http://dav.sourceforge.net/*, 2004.
21. Neon: *an HTTP and WebDAV client library, with a C interface, http://www.webdav.org/neon/*, 2004.
22. Collins, P. *PerlDAV: A WebDAV client library for Perl5, http://www.webdav.org/perldav/*, 2001.
23. JMOL *Molecular Viewer, http://jmol.sourceforge.net/*, 2005.

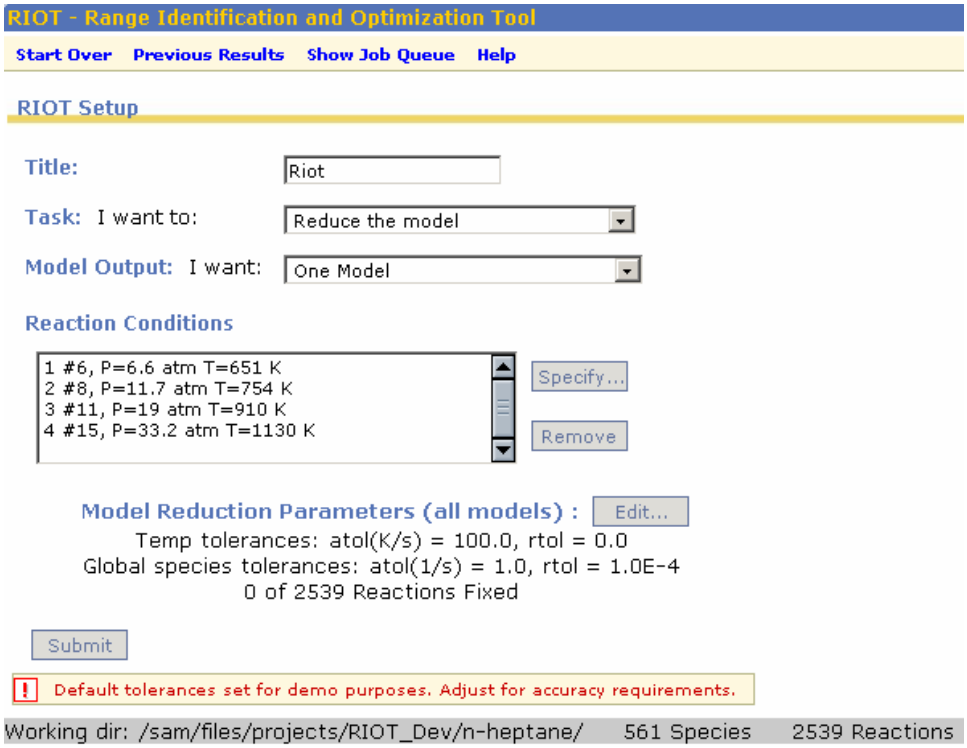

Figure 1: View of RIOT portal.

Reactions

Index	Reaction	A	beta	E
1	ch3+h(+M)<=>ch4(+M)	2.1380E+15	-0.4000	0.000
7	co+oh<=>co2+h	1.4000E+05	1.950	-1347.
8	h+o2<=>o+oh	1.9700E+14	0.000	1.6540E+04
10	o+h2o<=>2oh	2.9700E+06	2.020	1.3400E+04
11	oh+h2<=>h+h2o	2.1600E+08	1.510	3430.
14	c2h4+o<=>ch3+hco	1.0200E+07	1.880	179.0
15	h+c2h4(+M)<=>c2h5(+M)	1.0810E+12	0.4500	1822.
16	ch3oh(+M)<=>ch3+oh(+M)	1.9000E+16	0.000	9.1730E+04
22	ch3+ho2<=>ch3o+oh	1.1000E+13	0.000	0.000

Figure 2: Table view of partial listing of reduced mechanism obtained by automatic translation in portal.

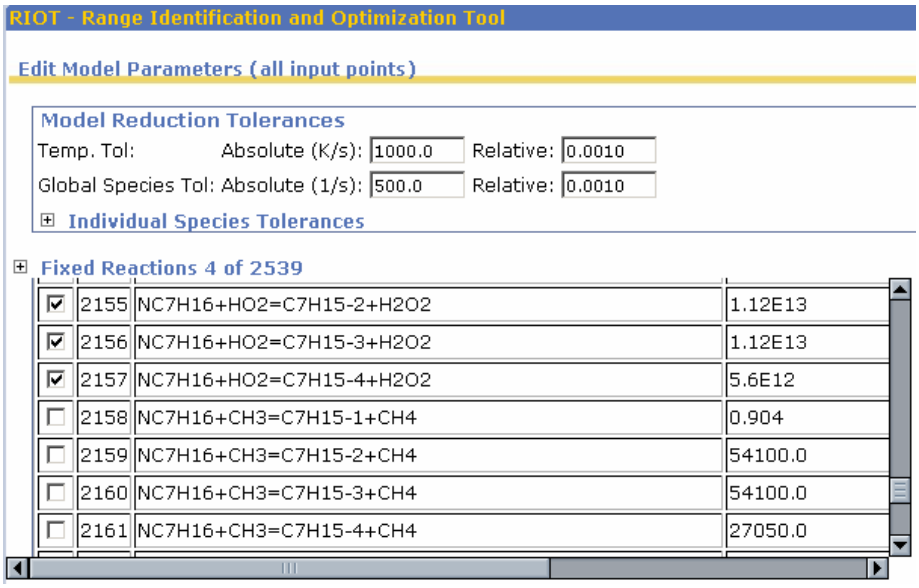

Figure 3: Menu allowing user-specified reactions which should remain in the reduced model.

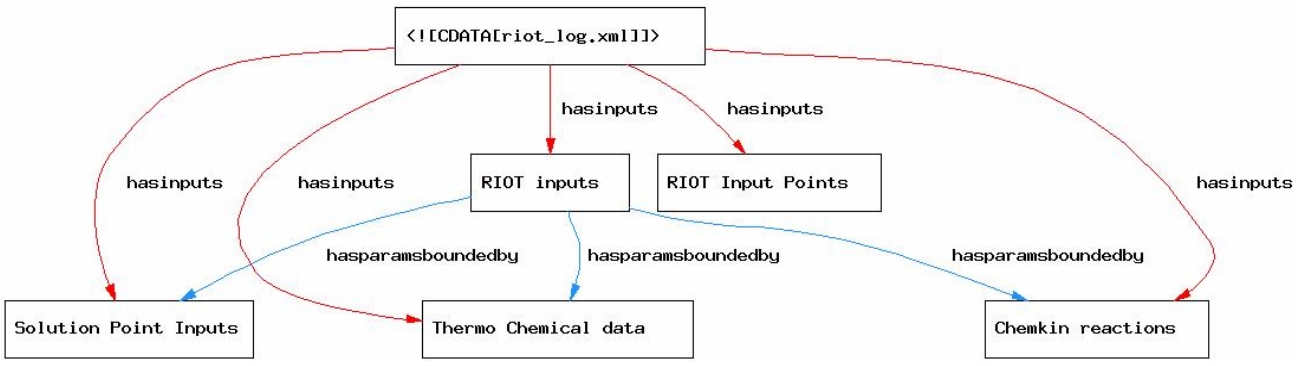

Figure 4: Graph of the pedigree of the reduced mechanism XML file.